www.sciencemag.org/content/355/6332/1436/suppl/DC1

Supplementary Materials for

On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria

Rochelle M. Soo, James Hemp, Donovan H. Parks, Woodward W. Fischer,* Philip Hugenholtz*

*Corresponding author. Email: p.hugenholtz@uq.edu.au (P.H.); wfischer@caltech.edu (W.W.F.)

Published 31 March 2017, *Science* **355**, 1436 (2017)
DOI: 10.1126/science.aal3794

This PDF file includes:

Materials and Methods

Data reporting
No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Datasets of publicly available sequencing reads
Paired-end read sequences from the following sequencing projects were downloaded from the NCBI Sequence Read Archive (SRA): SRR636581; SRR944699; SRR845265; ERR636395; ERR525992; ERR525373; ERR209866; ERR209865; ERR209449; ERR209447; ERR209359; ERR636350; ERR209439; ERR525787; ERR209450; ERR209669; ERR209448; ERR209516; ERR209606; ERR209863; ERR209350; ERR209864; ERR209110; ERR525875; ERR688535 and ERR209608. Paired-end reads were also downloaded from sequencing projects available through the Integrated Microbial Genomes (IMG) database *(26)*: metagenomes 3300000574; 3300003764; 3300003765, and composite assembled genomes 2582580516 and 2582580591.

Metagenome assembly and genome binning
Reads were quality trimmed using Trimmomatic (v.0.32) *(27)* with a leading/trailing value (Q3) and sliding window threshold (Q15). Trimmed reads were merged using BBMerge (v5.5) with default settings (https://sourceforge.net/projects/bbmap/). Trimmed paired-end reads were assembled using the CLC Genomics Workbench 7.0 (CLCBio, Qiagen) *de novo* assembly algorithm, using a k-mer size of 63 and a minimum scaffold length of 1000 bp. Sequencing reads were mapped to assembled scaffolds using BamM (v1.3.8) (http://ecogenomics.github.io/BamM/) which makes use of BWA (v0.7.12) *(28)* for read mapping. Population genome bins were recovered using MetaBAT (v.0.25.4) *(29)* with all settings (sensitive, specific, very sensitive and very specific) and the most complete bins based on the presence/absence of 104 bacterial single copy marker genes as defined by CheckM (v.1.0.3) *(30)* were used for further analysis. RefineM (v.0.0.3) (https://github.com/dparks1134/RefineM) was used to improve completeness and check for contamination of the population genomes.

Concatenated protein tree
A concatenated protein tree of 120 single copy marker genes (**Table S2**) was constructed from 13,843 genomes obtained from NCBI RefSeq (v75, March 2016), previously assembled *Melainabacteria* genomes, and the population genomes extracted from the SRA and IMG in the present study. Genes were aligned with HMMER v3.1b1 *(31)* and a maximum-likelihood tree was inferred with FastTree (v2.1.7, WAG + GAMMA model, other parameters set to default) *(32)*. Bootstrapping (100 times) was performed on a subset of the genomes consisting of all Cyanobacteria, Synergistetes, Chloroflexi and Armatimonadetes. Inferred trees were imported into ARB *(33)* for visualization and exported to Adobe Illustrator for figure production. Putative representatives of the *Melainabacteria* and *Sericytochromatia* were identified by their relative position in the tree, and corroborated where possible by 16S rRNA gene sequences present in the population genomes.

Single protein trees
Complex III and IV genes were identified in the *Melainabacteria* and *Sericytochromatia* genomes using the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4) *(25)*, which uses Prodigal v2.50 *(34)* to identify protein-coding genes. The protein sequences encoded by these genes and functionally characterized representatives from UniProt were used as seeds to identify homologues in the NCBI protein database using Mingle (v 0.0.7) (https://github.com/Ecogenomics/mingle). Briefly, matches with an *E*-value $\geq 1e^{-5}$, an amino acid identity $\geq 30\%$, and an alignment length $\geq 50\%$ were considered to be homologous. Sets of homologous proteins were aligned using MAFFT

(35) (v.7.221) and alignments filtered by excluding columns with <50% representation by homologous amino acids. Phylogenetic inference was performed on the filtered alignments using FastTree v2.1.7 *(32)* under the WAG + GAMMA model and bootstrapped 100 times using non-parametric bootstrapping. Inferred trees were imported into ARB for visualization and exported to Adobe Illustrator for figure production.

Channel prediction

Structural homology models for A-family oxygen reductases were built using SwissModel via the Swiss-PdbViewer interface. Proton channels were predicted comparing sequences alignments with the structural models.

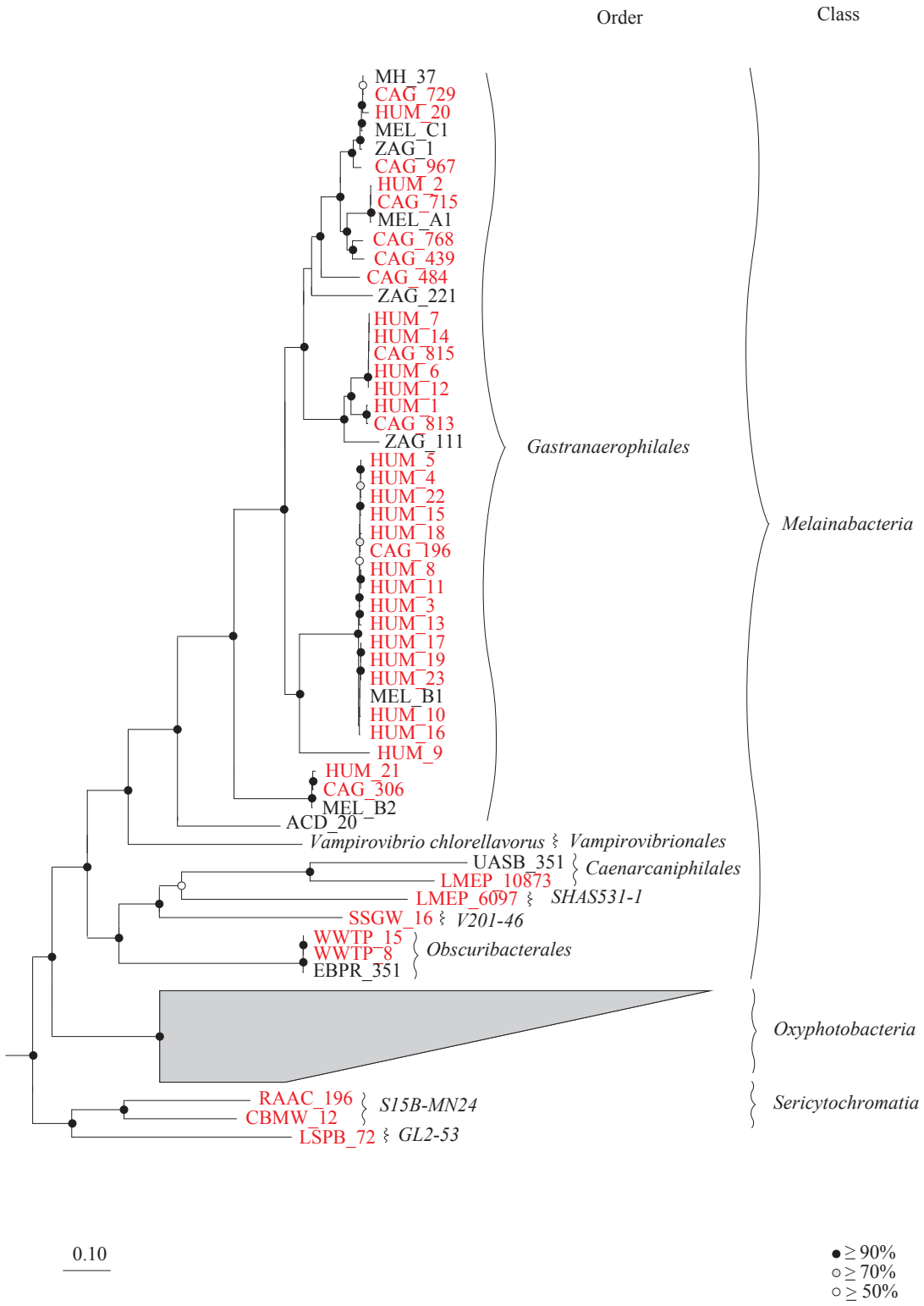


Fig. S1. Concatenated protein tree of the phylum Cyanobacteria. A maximum likelihood tree of the phylum Cyanobacteria based on the concatenated alignment of 120 phylogenetically conserved proteins (**Table S2**). Bootstrap resampling analyses (100 times) with maximum likelihood was performed with FastTree *(32)*. Taxa in red represent *Melainabacteria* and *Sericytochromatia* genomes obtained in the present study. Black circles represent interior nodes with ≥90% bootstrap support, grey circles ≥70% bootstrap support and white circles ≥50% bootstrap support. Order and class-level affiliations of the taxa are shown to the right of the figure. Outgroups (not shown) used for the analysis belong to the phyla Synergistetes, Chloroflexi and Armatimonadetes.

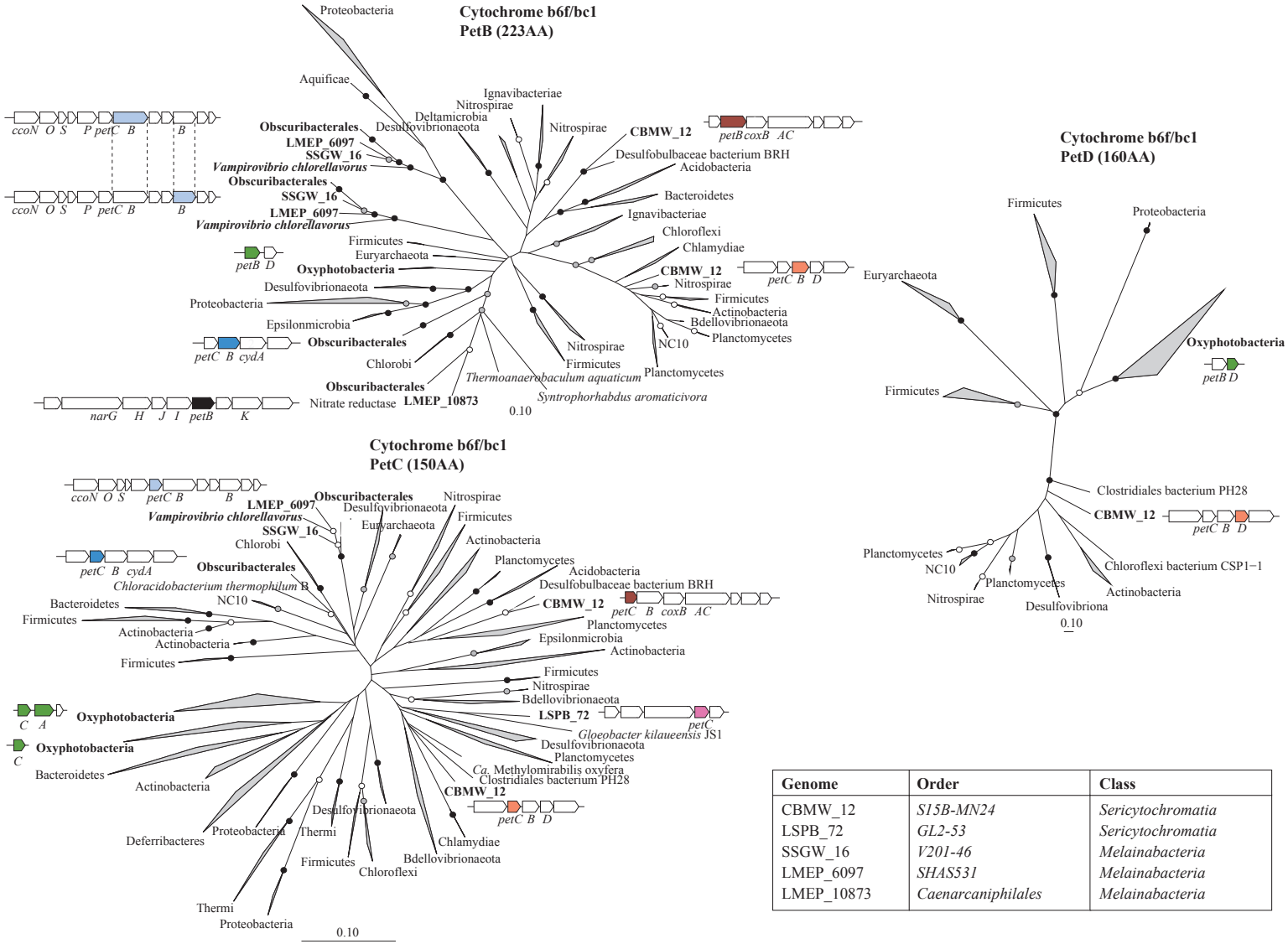
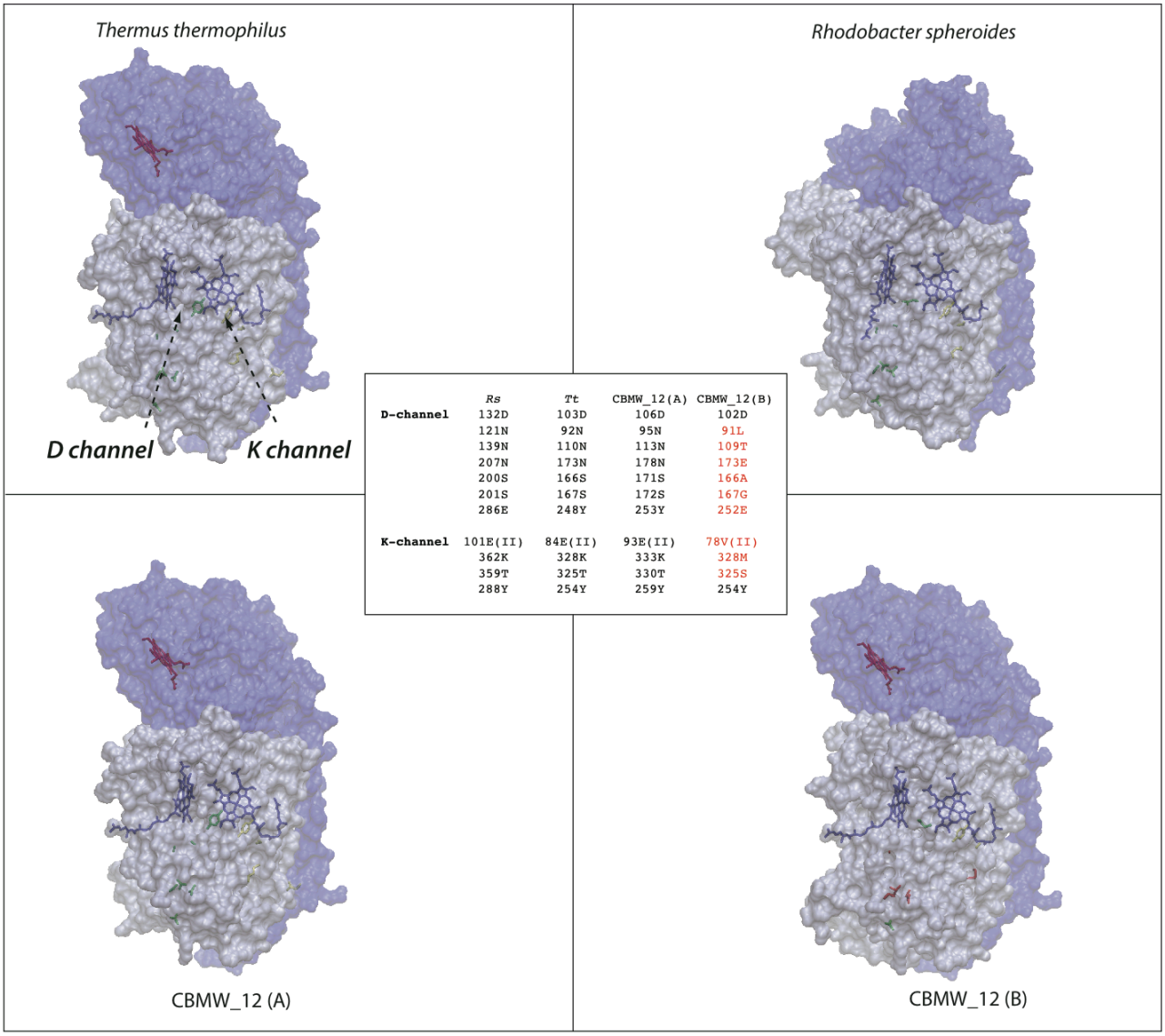
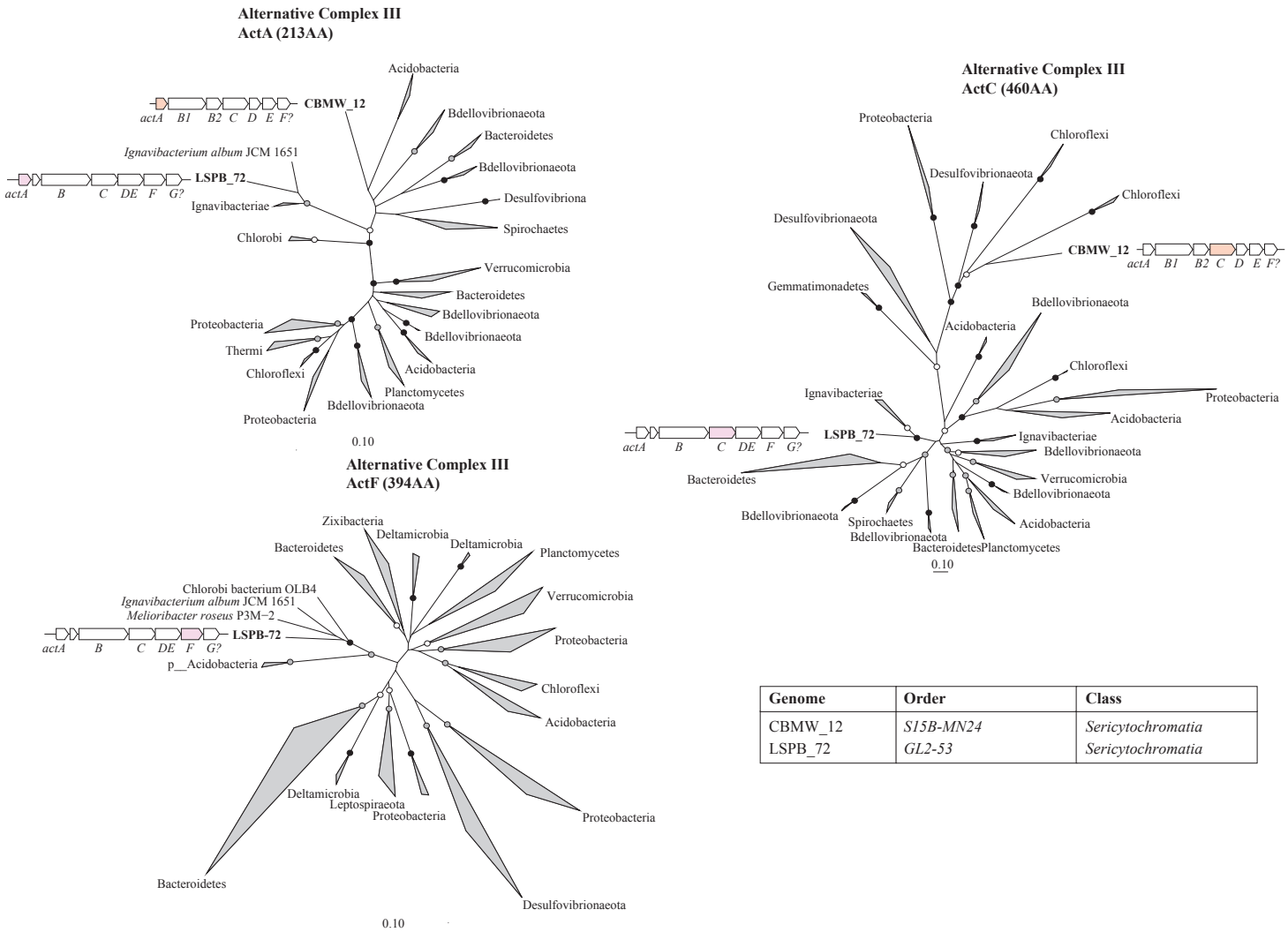


Fig. S2. Cytochrome *bc* complex protein trees. Maximum-likelihood trees of PetB (cytochrome b_6/b), PetC (cytochrome b_6f complex iron-sulfur subunit/cytochrome c_1) and PetD (cytochrome b_6f complex subunit IV) proteins with 100 bootstrap resamplings. Proteins belonging to Cyanobacteria are bolded in each tree and their gene neighborhood shown nearby, with the ortholog of interest shown in color. Bootstrap support for interior is indicated as per **Fig. S1**. Order and class-level assignments of individual cyanobacterial taxa is shown in the legend at the bottom right of the figure.



	Rs	*Tt*	CBMW_12(A)	CBMW_12(B)
D-channel	132D	103D	106D	102D
	121N	92N	95N	91L
	139N	110N	113N	109T
	207N	173N	178N	173E
	200S	166S	171S	166A
	201S	167S	172S	167G
	286E	248Y	253Y	252E
K-channel	101E(II)	84E(II)	93E(II)	78V(II)
	362K	328K	333K	328M
	359T	325T	330T	325S
	288Y	254Y	259Y	254Y

Fig. S3. Structure of the proton channels of the A-family oxygen reductases in *Sericytochromatia.*
D-channels are shown in green, and K-channels in yellow. The CBMW_12 (A) oxygen reductase has a Y at the top of the D-channel, similar to that from *Thermus thermophilus*. Both of its channels are completely conserved. The CBMW_12 (B) oxygen reductase has a glutamic acid at the top of the D-channel. Both D- and K-channels are highly modified (residues in red). The specific mutations imply that while the active-site receives protons from the cytoplasm, it is decoupled from proton pumping, and thus maximally conserves $1H^+$ per e^-.



Genome	Order	Class
CBMW_12	*S15B-MN24*	*Sericytochromatia*
LSPB_72	*GL2-53*	*Sericytochromatia*

Fig. S4. Alternative complex III protein trees.
Maximum-likelihood trees of ActA (alternative complex III molybdopterin oxidoreductase pentaheme c subunit), ActC (alternative complex III, protein C) and ActF (alternative complex III, protein F) proteins with 100 bootstrap resamplings. Proteins belonging to Cyanobacteria are bolded in each tree and their gene neighborhood shown nearby, with the ortholog of interest shown in color. Bootstrap support for interior is indicated as per **Fig. S1**. Order and class-level assignments of individual cyanobacterial taxa is shown in the legend at the bottom right of the figure.

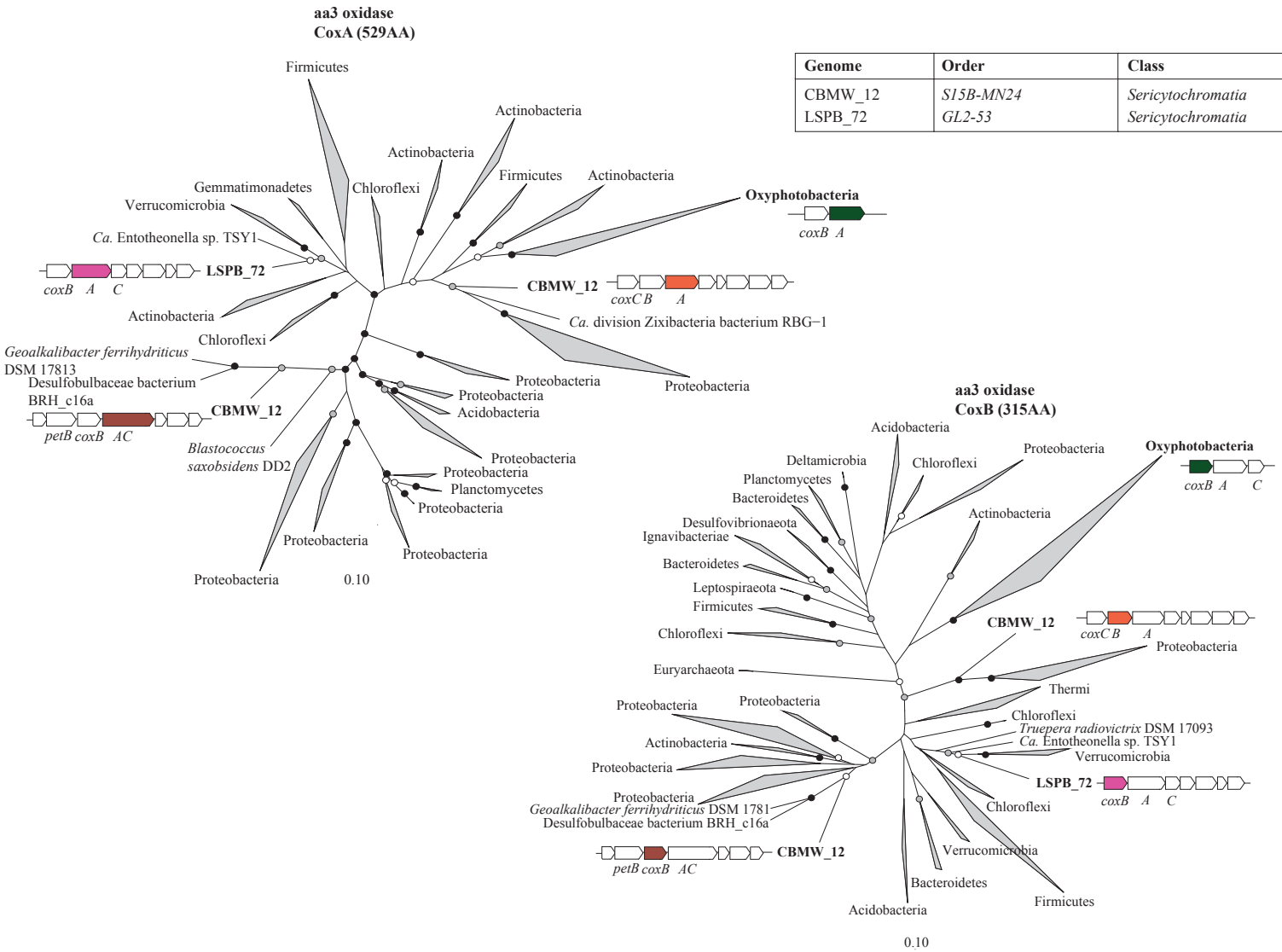
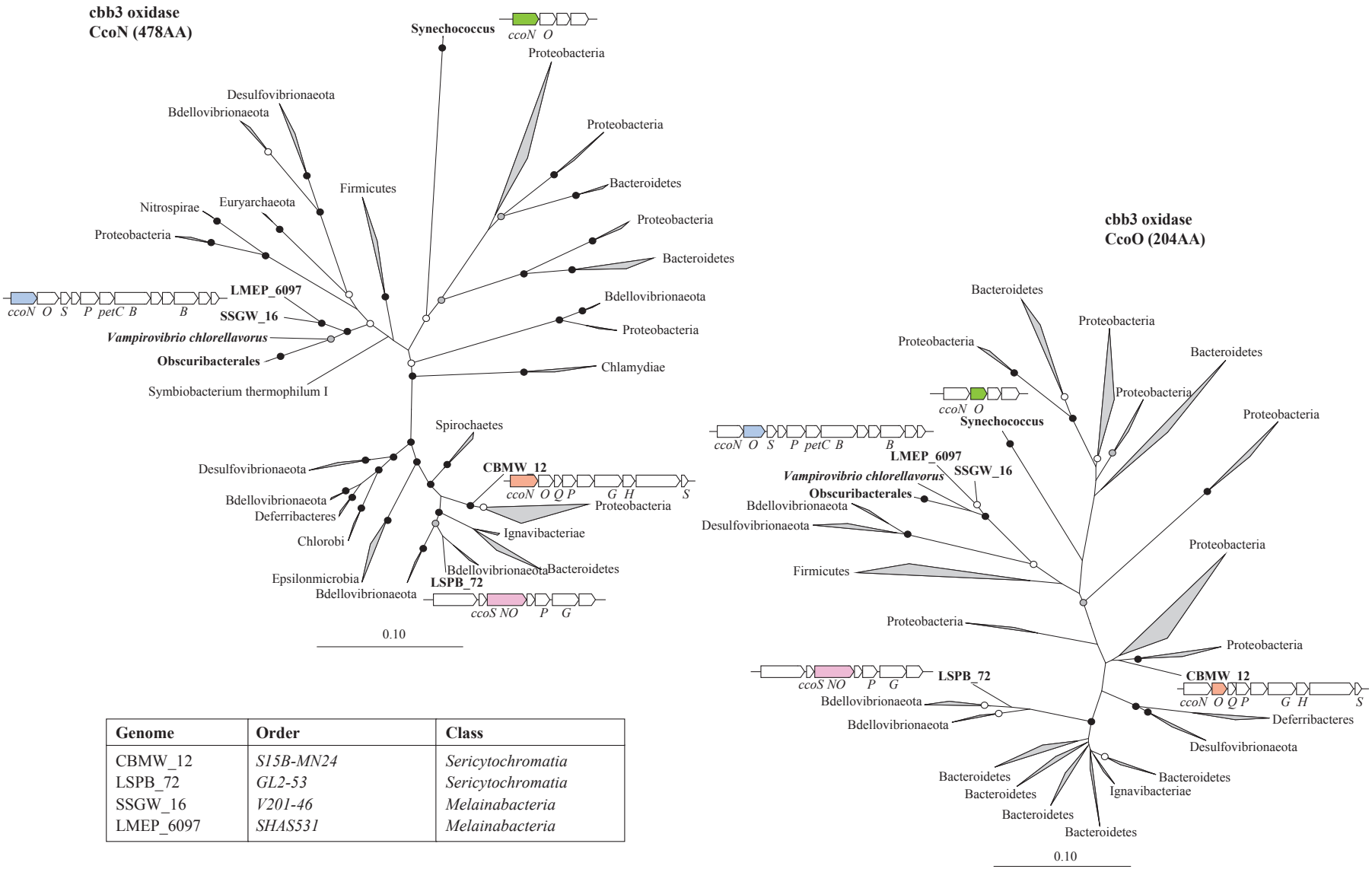


Fig. S5. HCO A-family oxygen reductase protein trees.
Maximum-likelihood trees of CoxA (cytochrome c oxidase subunit I) and CoxB (cytochrome c oxidase subunit II) proteins with 100 bootstrap resamplings. Proteins belonging to Cyanobacteria are bolded in each tree and their gene neighborhood shown nearby, with the ortholog of interest shown in color. Bootstrap support for interior is indicated as per **Fig. S1**. Order and class-level assignments of individual cyanobacterial taxa is shown in the legend at the top right of the figure.



Genome	Order	Class
CBMW_12	*S15B-MN24*	*Sericytochromatia*
LSPB_72	*GL2-53*	*Sericytochromatia*
SSGW_16	*V201-46*	*Melainabacteria*
LMEP_6097	*SHAS531*	*Melainabacteria*

Fig. S6. HCO C-family oxygen reductase protein trees.
Maximum-likelihood trees of CcoN (cytochrome c oxidase, *cbb*$_3$-type, subunit I) and CcoO (cytochrome c oxidase, *cbb*$_3$-type, subunit II) proteins with 100 bootstrap resamplings. Proteins belonging to Cyanobacteria are bolded in each tree and their gene neighborhood shown nearby, with the ortholog of interest shown in color. Bootstrap support for interior is indicated as per **Fig. S1**. Order and class-level assignments of individual cyanobacterial taxa is shown in the legend at the bottom left of the figure.

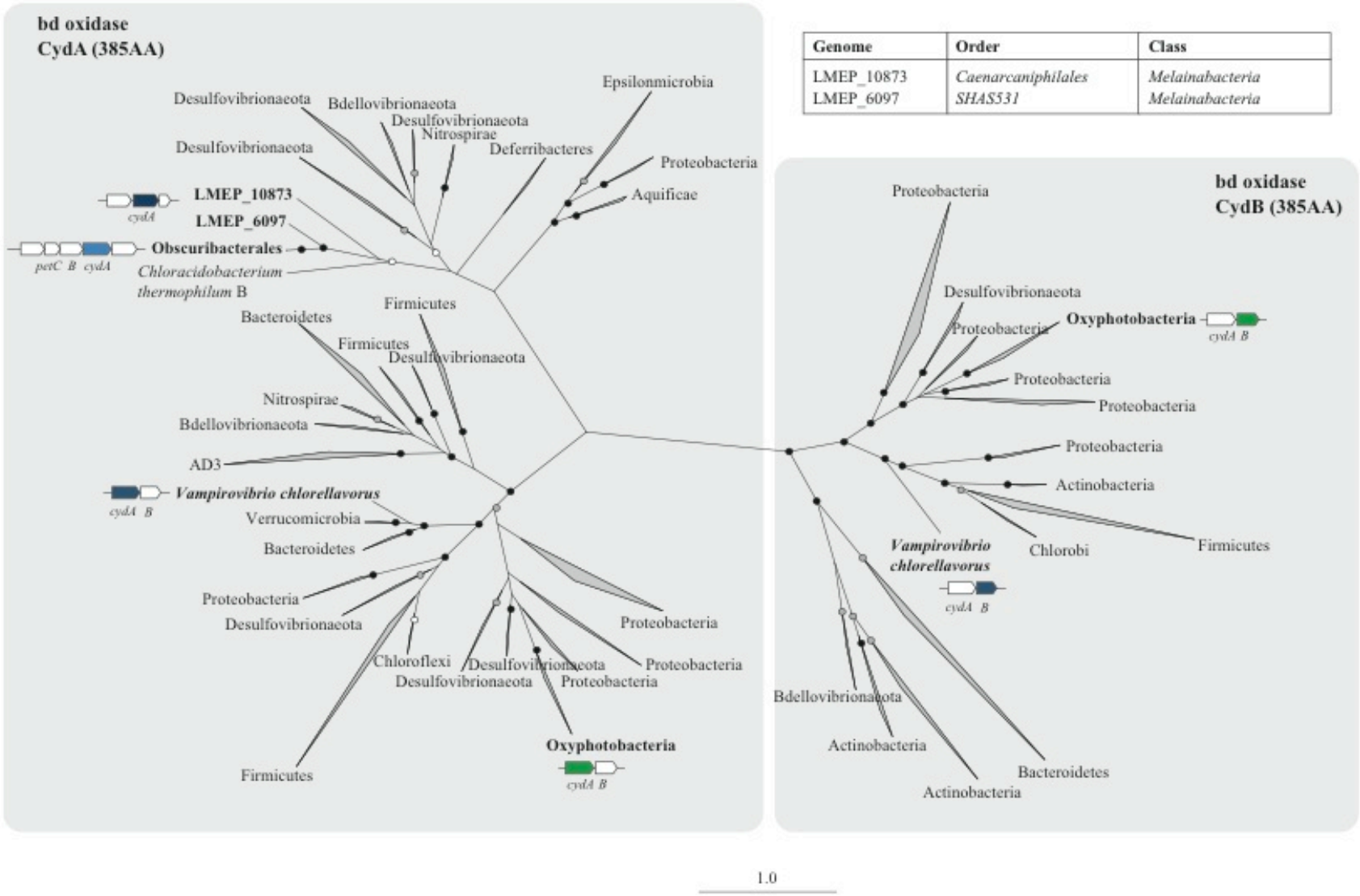


Fig. S7. Cytochrome *bd* oxidase protein tree.
Maximum-likelihood tree of CydA (cytochrome *bd* oxidase subunit I) and CydB (cytochrome *bd* oxidase subunit II) proteins with 100 bootstrap resamplings. Proteins belonging to Cyanobacteria are bolded in each tree and their gene neighborhood shown nearby, with the ortholog of interest shown in color. Bootstrap support for interior is indicated as per **Fig. S1**. Order and class-level assignments of individual cyanobacterial taxa is shown in the legend at the top right of the figure.

Table S1. Summary statistics of the *Melainabacteria* and *Sericytochromatia* genomes used in the study

Genome	Class[1]	Order[2]	Genome size (Mbp)	Number of Scaffolds	Number of Genes	16S[3]	CP[4] (%)	CT[5] (%)	Bioproject number/SRA metagenome acc.	Ref[6]
CBMW_12	Seri	S15B-MN24	3.9	59	3632	F	94.8	0	PRJNA348150	PS
RAAC_196	Seri	S15B-MN24	3.3	270	3373	N	91.8	1.7	PRJNA348151	PS
LSPB_72	Seri	GL2-53	4.5	563	4772	P	80.3	1.7	PRJNA348152	PS
MEL_A1	Mela	Gastranaero	1.9	1	1879	F	100	0	PRJNA321218	*36*
MEL_B1	Mela	Gastranaero	2.3	21	2269	F	98.3	0	**MEL.B1	*36*
MEL_B2	Mela	Gastranaero	2.3	26	2262	F	100	0	**MEL.B2	*36*
MEL_C1	Mela	Gastranaero	2.1	4	2162	F	100	0	**MEL.C1	*36*
ACD_20	Mela	Gastranaero	2.7	104	2565	N	100	3.5	PRJNA114691	*36*
ZAG_111	Mela	Gastranaero	2.2	65	2313	F	88.9	3.9	PRJNA347484	*1*
ZAG_221	Mela	Gastranaero	1.8	14	1838	N	89.5	0.9	PRJNA347484	*1*
MH_37	Mela	Gastranaero	2.1	157	2402	P	88.9	0	PRJNA347487	*1*
ZAG_1	Mela	Gastranaero	2	322	2194	N	82.9	0.9	PRJNA347484	*1*
HUM_1	Mela	Gastranaero	1.9	8	1845	N	100	0	PRJNA348149	PS
HUM_2	Mela	Gastranaero	1.9	49	1924	N	100	0	PRJNA348149	PS
HUM_3	Mela	Gastranaero	2.2	48	2175	F	100	0	PRJNA348149	PS
HUM_4	Mela	Gastranaero	2.3	39	2323	N	100	0	PRJNA348149	PS
HUM_5	Mela	Gastranaero	2.3	34	2313	F	100	0	PRJNA348149	PS
HUM_6	Mela	Gastranaero	1.8	52	1861	N	100	0	PRJNA348149	PS
HUM_7	Mela	Gastranaero	1.9	48	1955	N	99.1	0	PRJNA348149	PS
HUM_8	Mela	Gastranaero	2.1	47	2191	N	99.1	2.5	PRJNA348149	PS
HUM_9	Mela	Gastranaero	1.9	20	1692	N	98.3	1.7	PRJNA348149	PS
HUM_10	Mela	Gastranaero	2.5	85	2545	F	98.3	0	PRJNA348149	PS
HUM_11	Mela	Gastranaero	2.3	113	2310	N	97.4	0	PRJNA348149	PS
HUM_12	Mela	Gastranaero	1.9	56	1880	F	97.4	0	PRJNA348149	PS
HUM_13	Mela	Gastranaero	2.1	83	2141	N	96.6	1.2	PRJNA348149	PS
HUM_14	Mela	Gastranaero	1.8	60	1783	N	96.6	0.9	PRJNA348149	PS
HUM_15	Mela	Gastranaero	2.1	49	2146	P	94.8	0	PRJNA348149	PS
HUM_16	Mela	Gastranaero	2	44	2030	N	94.0	0.9	PRJNA348149	PS
HUM_17	Mela	Gastranaero	2.3	259	2254	N	93.1	4.0	PRJNA348149	PS
HUM_18	Mela	Gastranaero	2.7	116	2891	N	93.0	1.7	PRJNA348149	PS
HUM_19	Mela	Gastranaero	2.1	151	2140	N	91.0	3.6	PRJNA348149	PS
HUM_20	Mela	Gastranaero	1.8	237	1886	N	89.6	0	PRJNA348149	PS
HUM_21	Mela	Gastranaero	1.9	225	1975	N	85.2	0.3	PRJNA348149	PS
HUM_22	Mela	Gastranaero	1.9	97	1844	N	87.9	0	PRJNA348149	PS
HUM_23	Mela	Gastranaero	1.9	26	1953	N	81.0	0	PRJNA348149	PS
CAG_196	Mela	Gastranaero	2.1	53	2148	F	100	0	PRJNA221948	PS
CAG_306	Mela	Gastranaero	2.2	73	2197	F	96.6	0	PRJNA222071	PS
CAG_439	Mela	Gastranaero	2.2	90	2367	N	100	0	PRJNA222187	PS
CAG_484	Mela	Gastranaero	2.2	62	2263	P	99.1	0	PRJNA221894	PS

CAG_715	Mela	Gastranaero	1.9	40	1902	N	100	0	PRJNA221754	PS
CAG_729	Mela	Gastranaero	2	48	2063	N	98.3	0	PRJNA222200	PS
CAG_768	Mela	Gastranaero	2	18	2034	N	100	0	PRJNA222178	PS
CAG_815	Mela	Gastranaero	1.8	32	1861	N	100	0	PRJNA222202	PS
CAG_813	Mela	Gastranaero	1.9	76	1974	N	98.3	0	PRJNA222206	PS
CAG_967	Mela	Gastranaero	2	42	2081	N	98.3	0	PRJNA221908	PS
EBPR_351	Mela	Obscuribact	5.5	8	4655	F	97.4	2.3	PRJNA347481	*1*
WWTP_8	Mela	Obscuribact	3.4	5	2855	F	87.9	0.9	*3300003765	PS
WWTP_15	Mela	Obscuribact	4.8	26	4042	F	98.3	0.9	*3300003764	PS
SSGW_16	Mela	V201-46	2.3	46	2062	F	100	0	*3300000574	PS
LMEP_10873	Mela	Caenarcani	2.2	154	2116	F	87.8	3.5	PRJNA337808	PS
UASB_351	Mela	Caenarcani	1.8	67	1913	P	84.6	0.9	PRJNA347483	*1*
LMEP_6097	Mela	SHAS531	1.9	25	1819	F	93.9	0	PRJNA337808	PS
Vampirovibrio chlorellavorus	Mela	Vampirovibrio	3	28	2844	F	100	1.7	PRJNA278896	*37*

[1]Class = Seri (*Sericytochromatia*), Mela (*Melainabacteria*)
[2]Order = Gastranaero (*Gastranaerophilales*), Obscuribact (*Obscuribacterales*) Caenarcani (*Caenarcaniphilales*), Vampirovibrio (*Vampirovibrionales*)
[3]16S = P (Partial) <1300bp, F (Full) >1300bp, N (None)
[4]CP = Estimated completeness based on the presence of 104 single copy genes *(30)*
[5]CT = Estimated contamination based on the presence of more than one single copy gene *(30)*
[6]Ref = PS (Present study)
* IMG accession number/genome not deposited in Genbank
** ggkBase accession name/genome not deposited in Genbank

Table S2. A list of the 120 bacterial marker genes used for the concatenated protein tree inference

Marker ID	Name	Description	Length (aa)
PF02576.12	DUF150	Uncharacterised BCR, YhbC family COG0779	141
PF01025.14	GrpE	GrpE	166
PF03726.9	PNPase	Polyribonucleotide nucleotidyltransferase, RNA binding domain	83
PF00466.15	Ribosomal_L10	Ribosomal protein L10	100
PF00410.14	Ribosomal_S8	Ribosomal protein S8	129
PF00380.14	Ribosomal_S9	Ribosomal protein S9/S16	121
TIGR00006	TIGR00006	16S rRNA (cytosine(1402)-N(4))-methyltransferase	310
TIGR00019	prfA	peptide chain release factor 1	361
TIGR00020	prfB	peptide chain release factor 2	365
TIGR00029	S20	ribosomal protein bS20	87
TIGR00043	TIGR00043	rRNA maturation RNase YbeY	111
TIGR00054	TIGR00054	RIP metalloprotease RseP	421
TIGR00059	L17	ribosomal protein bL17	112
TIGR00061	L21	ribosomal protein bL21	101
TIGR00064	ftsY	signal recognition particle-docking protein FtsY	279
TIGR00065	ftsZ	cell division protein FtsZ	353
TIGR00082	rbfA	ribosome-binding factor A	115
TIGR00083	ribF	riboflavin biosynthesis protein RibF	290
TIGR00084	ruvA	Holliday junction DNA helicase RuvA	192
TIGR00086	smpB	SsrA-binding protein	144
TIGR00088	trmD	tRNA (guanine(37)-N(1))-methyltransferase	233
TIGR00090	rsfS_iojap_ybeB	ribosome silencing factor	99
TIGR00092	TIGR00092	GTP-binding protein YchF	368
TIGR00095	TIGR00095	16S rRNA (guanine(966)-N(2))-methyltransferase RsmD	194
TIGR00115	tig	trigger factor	410
TIGR00116	tsf	translation elongation factor Ts	293
TIGR00138	rsmG_gidB	16S rRNA (guanine(527)-N(7))-methyltransferase RsmG	183
TIGR00158	L9	ribosomal protein bL9	148
TIGR00166	S6	ribosomal protein bS6	95
TIGR00168	infC	translation initiation factor IF-3	165
TIGR00186	rRNA_methyl_3	RNA methyltransferase, TrmH family, group 3	240
TIGR00194	uvrC	excinuclease ABC subunit C	574
TIGR00250	RNAse_H_YqgF	putative transcription antitermination factor YqgF	130
TIGR00337	PyrG	CTP synthase	526
TIGR00344	alaS	alanine--tRNA ligase	847
TIGR00362	DnaA	chromosomal replication initiator protein DnaA	437
TIGR00382	clpX	ATP-dependent Clp protease, ATP-binding subunit ClpX	414
TIGR00392	ileS	isoleucine--tRNA ligase	861
TIGR00396	leuS_bact	leucine--tRNA ligase	843
TIGR00398	metG	methionine--tRNA ligase	530
TIGR00414	serS	serine--tRNA ligase	418
TIGR00416	sms	DNA repair protein RadA	454
TIGR00420	trmU	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	351
TIGR00431	TruB	tRNA pseudouridine(55) synthase	210
TIGR00435	cysS	cysteine--tRNA ligase	466
TIGR00436	era	GTP-binding protein Era	270
TIGR00442	hisS	histidine--tRNA ligase	406
TIGR00445	mraY	phospho-N-acetylmuramoyl-pentapeptide-transferase	321
TIGR00456	argS	arginine--tRNA ligase	569
TIGR00459	aspS_bact	aspartate--tRNA ligase	586
TIGR00460	fmt	methionyl-tRNA formyltransferase	315
TIGR00468	pheS	phenylalanine--tRNA ligase, alpha subunit	324
TIGR00472	pheT_bact	phenylalanine--tRNA ligase, beta subunit	798
TIGR00487	IF-2	translation initiation factor IF-2	587
TIGR00496	frr	ribosome recycling factor	176
TIGR00539	hemN_rel	putative oxygen-independent coproporphyrinogen III oxidase	361
TIGR00580	mfd	transcription-repair coupling factor	923
TIGR00593	pola	DNA polymerase I	890
TIGR00615	recR	recombination protein RecR	196
TIGR00631	uvrb	excinuclease ABC subunit B	658
TIGR00634	recN	DNA repair protein RecN	563
TIGR00635	ruvB	Holliday junction DNA helicase RuvB	305
TIGR00643	recG	ATP-dependent DNA helicase RecG	629
TIGR00663	dnan	DNA polymerase III, beta subunit	367
TIGR00717	rpsA	ribosomal protein bS1	516
TIGR00755	ksgA	ribosomal RNA small subunit methyltransferase A	256
TIGR00810	secG	preprotein translocase, SecG subunit	73
TIGR00922	nusG	transcription termination/antitermination factor NusG	172

TIGR00928	purB	adenylosuccinate lyase	436
TIGR00959	ffh	signal recognition particle protein	428
TIGR00963	secA	preprotein translocase, SecA subunit	787
TIGR00964	secE_bact	preprotein translocase, SecE subunit	57
TIGR00967	3a0501s007	preprotein translocase, SecY subunit	414
TIGR01009	rpsC_bact	ribosomal protein uS3	212
TIGR01011	rpsB_bact	ribosomal protein uS2	225
TIGR01017	rpsD_bact	ribosomal protein uS4	200
TIGR01021	rpsE_bact	ribosomal protein uS5	156
TIGR01029	rpsG_bact	ribosomal protein uS7	154
TIGR01032	rplT_bact	ribosomal protein bL20	114
TIGR01039	atpD	ATP synthase F1, beta subunit	462
TIGR01044	rplV_bact	ribosomal protein uL22	103
TIGR01059	gyrB	DNA gyrase, B subunit	639
TIGR01063	gyrA	DNA gyrase, A subunit	800
TIGR01066	rplM_bact	ribosomal protein uL13	141
TIGR01071	rplO_bact	ribosomal protein uL15	144
TIGR01079	rplX_bact	ribosomal protein uL24	104
TIGR01082	murC	UDP-N-acetylmuramate--L-alanine ligase	449
TIGR01087	murD	UDP-N-acetylmuramoylalanine--D-glutamate ligase	441
TIGR01128	holA	DNA polymerase III, delta subunit	314
TIGR01146	ATPsyn_F1gamma	ATP synthase F1, gamma subunit	286
TIGR01164	rplP_bact	ribosomal protein uL16	126
TIGR01169	rplA_bact	ribosomal protein uL1	227
TIGR01171	rplB_bact	ribosomal protein uL2	275
TIGR01302	IMP_dehydrog	inosine-5'-monophosphate dehydrogenase	450
TIGR01391	dnaG	DNA primase	414
TIGR01393	lepA	elongation factor 4	595
TIGR01394	TypA_BipA	GTP-binding protein TypA/BipA	594
TIGR01510	coaD_prev_kdtB	pantetheine-phosphate adenylyltransferase	155
TIGR01632	L11_bact	ribosomal protein uL11	140
TIGR01951	nusB	transcription antitermination factor NusB	131
TIGR01953	NusA	transcription termination factor NusA	340
TIGR02012	tigrfam_recA	protein RecA	321
TIGR02013	rpoB	DNA-directed RNA polymerase, beta subunit	1238
TIGR02027	rpoA	DNA-directed RNA polymerase, alpha subunit	298
TIGR02075	pyrH_bact	UMP kinase	233
TIGR02191	RNaseIII	ribonuclease III	219
TIGR02273	16S_RimM	16S rRNA processing protein RimM	166
TIGR02350	prok_dnaK	chaperone protein DnaK	596
TIGR02386	rpoC_TIGR	DNA-directed RNA polymerase, beta' subunit	1147
TIGR02397	dnaX_nterm	DNA polymerase III, subunit gamma and tau	355
TIGR02432	lysidine_TilS_N	tRNA(Ile)-lysidine synthetase	189
TIGR02729	Obg_CgtA	Obg family GTPase CgtA	329
TIGR03263	guanyl_kin	guanylate kinase	180
TIGR03594	GTPase_EngA	ribosome-associated GTPase EngA	432
TIGR03625	L3_bact	50S ribosomal protein uL3	202
TIGR03632	uS11_bact	ribosomal protein uS11	117
TIGR03654	L6_bact	ribosomal protein uL6	175
TIGR03723	T6A_TsaD_YgjD	tRNA threonylcarbamoyl adenosine modification protein TsaD	314
TIGR03725	T6A_YeaZ	tRNA threonylcarbamoyl adenosine modification protein YeaZ	212
TIGR03953	rplD_bact	50S ribosomal protein uL4	188

References and Notes

1. R. M. Soo, C. T. Skennerton, Y. Sekiguchi, M. Imelfort, S. J. Paech, P. G. Dennis, J. A. Steen, D. H. Parks, G. W. Tyson, P. Hugenholtz, An expanded genomic representation of the phylum cyanobacteria. *Genome Biol. Evol.* **6**, 1031–1045 (2014). doi:10.1093/gbe/evu073 Medline

2. P. M. Shih, J. Hemp, L. M. Ward, N. J. Matzke, W. W. Fischer, Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobiology* **15**, 19–29 (2017). doi:10.1111/gbi.12200 Medline

3. D. An, S. M. Caffrey, J. Soh, A. Agrawal, D. Brown, K. Budwill, X. Dong, P. F. Dunfield, J. Foght, L. M. Gieg, S. J. Hallam, N. W. Hanson, Z. He, T. R. Jack, J. Klassen, K. M. Konwar, E. Kuatsjah, C. Li, S. Larter, V. Leopatra, C. L. Nesbø, T. Oldenburg, A. P. Pagé, E. Ramos-Padron, F. F. Rochman, A. Saidi-Mehrabad, C. W. Sensen, P. Sipahimalani, Y. C. Song, S. Wilson, G. Wolbring, M.-L. Wong, G. Voordouw, Metagenomics of hydrocarbon resource environments indicates aerobic taxa and genes to be unexpectedly common. *Environ. Sci. Technol.* **47**, 10708–10717 (2013). doi:10.1021/es4020184 Medline

4. L. A. Hug, B. C. Thomas, C. T. Brown, K. R. Frischkorn, K. H. Williams, S. G. Tringe, J. F. Banfield, Aquifer environment selects for microbial species cohorts in sediment and groundwater. *ISME J.* **9**, 1846–1856 (2015). doi:10.1038/ismej.2015.2 Medline

5. H. B. Nielsen, M. Almeida, A. S. Juncker, S. Rasmussen, J. Li, S. Sunagawa, D. R. Plichta, L. Gautier, A. G. Pedersen, E. Le Chatelier, E. Pelletier, I. Bonde, T. Nielsen, C. Manichanh, M. Arumugam, J. M. Batto, M. B. Quintanilha Dos Santos, N. Blom, N. Borruel, K. S. Burgdorf, F. Boumezbeur, F. Casellas, J. Doré, P. Dworzynski, F. Guarner, T. Hansen, F. Hildebrand, R. S. Kaas, S. Kennedy, K. Kristiansen, J. R. Kultima, P. Léonard, F. Levenez, O. Lund, B. Moumen, D. Le Paslier, N. Pons, O. Pedersen, E. Prifti, J. Qin, J. Raes, S. Sørensen, J. Tap, S. Tims, D. W. Ussery, T. Yamada, P. Renault, T. Sicheritz-Ponten, P. Bork, J. Wang, S. Brunak, S. D. Ehrlich, MetaHIT Consortium, Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014). doi:10.1038/nbt.2939 Medline

6. W. W. Fischer, J. Hemp, J. E. Johnson, Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016). doi:10.1146/annurev-earth-060313-054810

7. M. F. Hohmann-Marriott, R. E. Blankenship, Evolution of photosynthesis. *Annu. Rev. Plant Biol.* **62**, 515–548 (2011). doi:10.1146/annurev-arplant-042110-103811 Medline

8. T. Cardona, A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynth. Res.* **126**, 111–134 (2015). doi:10.1007/s11120-014-0065-x Medline

9. F. L. Sousa, L. Shavit-Grievink, J. F. Allen, W. F. Martin, Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol. Evol.* **5**, 200–216 (2013). doi:10.1093/gbe/evs127 Medline

10. A. Y. Mulkidjanian, E. V. Koonin, K. S. Makarova, S. L. Mekhedov, A. Sorokin, Y. I. Wolf, A. Dufresne, F. Partensky, H. Burd, D. Kaznadzey, R. Haselkorn, M. Y. Galperin, The

cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13126–13131 (2006). doi:10.1073/pnas.0605709103 Medline

11. M. F. Yanyushin, M. C. del Rosario, D. C. Brune, R. E. Blankenship, New class of bacterial membrane oxidoreductases. *Biochemistry* **44**, 10037–10045 (2005). doi:10.1021/bi047267l Medline

12. P. N. Refojo, M. A. Ribeiro, F. Calisto, M. Teixeira, M. M. Pereira, Structural composition of alternative complex III: Variations on the same theme. *Biochim. Biophys. Acta* **1827**, 1378–1382 (2013). doi:10.1016/j.bbabio.2013.01.001

13. D. V. Dibrova, D. A. Cherepanov, M. Y. Galperin, V. P. Skulachev, A. Y. Mulkidjanian, Evolution of cytochrome *bc* complexes: From membrane-anchored dehydrogenases of ancient bacteria to triggers of apoptosis in vertebrates. *Biochim. Biophys. Acta* **1827**, 1407–1427 (2013). doi:10.1016/j.bbabio.2013.07.006 Medline

14. D. Stroebel, Y. Choquet, J.-L. Popot, D. Picot, An atypical haem in the cytochrome b_6f complex. *Nature* **426**, 413–418 (2003). doi:10.1038/nature02155 Medline

15. M. M. Pereira, M. Santana, M. Teixeira, A novel scenario for the evolution of haem–copper oxygen reductases. *Biochim. Biophys. Acta* **1505**, 185–208 (2001). doi:10.1016/S0005-2728(01)00169-4

16. J. Hemp, R. B. Gennis, Diversity of the heme-copper superfamily in archaea: Insights from genomics and structural modeling. *Results Probl. Cell Differ.* **45**, 1–31 (2008). doi:10.1007/400_2007_046 Medline

17. H. Han, J. Hemp, L. A. Pace, H. Ouyang, K. Ganesan, J. H. Roh, F. Daldal, S. R. Blanke, R. B. Gennis, Adaptation of aerobic respiration to low O_2 environments. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14109–14114 (2011). doi:10.1073/pnas.1018958108 Medline

18. V. B. Borisov, R. B. Gennis, J. Hemp, M. I. Verkhovsky, The cytochrome *bd* respiratory oxygen reductases. *Biochim. Biophys. Acta* **1807**, 1398–1413 (2011). doi:10.1016/j.bbabio.2011.06.016 Medline

19. G. Schmetterer, in *Cytochrome Complexes: Evolution, Structures, Energy Transduction, and Signaling*, A. W. Cramer, T. Kallas, Eds. (Springer, 2016), pp. 331–335.

20. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315 (2014). doi:10.1038/nature13068 Medline

21. A. D. Anbar, Y. Duan, T. W. Lyons, G. L. Arnold, B. Kendall, R. A. Creaser, A. J. Kaufman, G. W. Gordon, C. Scott, J. Garvin, R. Buick, A whiff of oxygen before the great oxidation event? *Science* **317**, 1903–1906 (2007). doi:10.1126/science.1140325 Medline

22. M. T. Rosing, R. Frei, U-rich Archaean seafloor sediments from Greenland – indications of >3700 Ma oxygenic photosynthesis. *Earth Planet. Sci. Lett.* **217**, 237–244 (2004). doi:10.1016/S0012-821X(03)00609-5

23. J. W. Schopf, B. M. Packer, Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* **237**, 70–73 (1987). doi:10.1126/science.11539686 Medline

24. J. E. Johnson, S. M. Webb, K. Thomas, S. Ono, J. L. Kirschvink, W. W. Fischer, Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11238–11243 (2013). doi:10.1073/pnas.1305530110 Medline

25. M. Huntemann, N. N. Ivanova, K. Mavromatis, H. J. Tripp, D. Paez-Espino, K. Palaniappan, E. Szeto, M. Pillay, I.-M. A. Chen, A. Pati, T. Nielsen, V. M. Markowitz, N. C. Kyrpides, The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand. Genomic Sci.* **10**, 86 (2015). doi:10.1186/s40793-015-0077-y Medline

26. V. M. Markowitz, I. M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N. N. Ivanova, N. C. Kyrpides, IMG: The Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012). doi:10.1093/nar/gkr1044 Medline

27. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). doi:10.1093/bioinformatics/btu170 Medline

28. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324 Medline

29. D. D. Kang, J. Froula, R. Egan, Z. Wang, MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015). doi:10.7717/peerj.1165 Medline

30. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, M. Check, Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015). doi:10.1101/gr.186072.114

31. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39** (suppl. 2), W29–37 (2011). doi:10.1093/nar/gkr367 Medline

32. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009). doi:10.1093/molbev/msp077 Medline

33. W. Ludwig, O. Strunk, R. Westram, L. Richter, H. Meier, B. A. Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, K. H. Schleifer, ARB: A software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004). doi:10.1093/nar/gkh293 Medline

34. D. Hyatt, G.-L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010). doi:10.1186/1471-2105-11-119 Medline

35. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). doi:10.1093/molbev/mst010 Medline

36. S. C. Di Rienzi, I. Sharon, K. C. Wrighton, O. Koren, L. A. Hug, B. C. Thomas, J. K. Goodrich, J. T. Bell, T. D. Spector, J. F. Banfield, R. E. Ley, The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *eLife* **2**, e01102 (2013). doi:10.7554/eLife.01102 Medline

37. R. M. Soo, B. J. Woodcroft, D. H. Parks, G. W. Tyson, P. Hugenholtz, Back from the dead; the curious tale of the predatory cyanobacterium *Vampirovibrio chlorellavorus*. *PeerJ* **3**, e968 (2015). doi:10.7717/peerj.968 Medline